NRL Report 9283

AD-A229 757

Explicit State Vector Representation for Heteroassociative Memories

JEFFREY K. UHLMANN

Battle Management Technology Branch
Information Technology Division

November 21, 1990

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
November 21, 1990

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Explicit State Vector Representation for Heteroassociative Memories

5. FUNDING NUMBERS
PE - 62702E
TA - 5625
WU - DN150-127

6. AUTHOR(S)
Jeffrey K. Uhlmann

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
4555 Overlook Ave.
Washington, DC 20375-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL Report 9283

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency
Naval Technology Program Office
1400 Wilson Blvd.
Arlington, VA 22209-2308

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A new vector representation is examined as an alternative to the bipolar form often used in associative memory models. This representation is shown to eliminate constraining symmetries through the introduction of a noncommutative correlation operator. It is also shown that this representation leads to a recursive formulation for a nonlinear associative memory with provable convergence properties. Generalizations using higher order correlations are described.

14. SUBJECT TERMS
Associative memories — Affine transformations
Neural networks — Higher order neural networks

15. NUMBER OF PAGES
9

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

EXPLICIT STATE VECTOR REPRESENTATION FOR HETEROASSOCIATIVE MEMORIES

INTRODUCTION

A critical issue in the analysis and development of associative memory (AM) models is the representation of information. Typically, information to be processed is presented in the form of a fixed-length vector of binary (two-state) variables. Clearly, the binary $(0,1)$ representation is sufficient; for a variety of reasons, however, the bipolar $(-1,+1)$ scheme is more commonly used. One reason in particular is that it avoids normalization considerations because every n-length bipolar vector has the same $\sqrt{n}$ magnitude. This report examines an alternative vector representation that maintains the advantages of the bipolar form while leading to more powerful, nonlinear AM formulations.

Given a set of bipolar vector pairs $\{(\mathbf{u}_1,\mathbf{v}_1)...(\mathbf{u}_n,\mathbf{v}_n)\}$, a bidirectional associative memory M can be constructed as a sum of outer products:

$$M = \sum_{i=1}^{n} \mathbf{u}_i^T \mathbf{v}_i. \tag{1}$$

This matrix is referred to as being bidirectionally associative [1] because for a learned pair $(\mathbf{u}_i,\mathbf{v}_i)$ it has the property that $\langle \mathbf{u}_i M\rangle = \mathbf{v}_i$ and $\langle \mathbf{v}_i M^T\rangle = \mathbf{u}_i$, where the $\langle\cdot\rangle$ function converts the elements of a vector to bipolar form (according to sign). In this single-pass model of an AM, associations are maintained as a matrix of first-order correlations between elements of the input vectors and elements of the output vectors. It is easy to establish that such a model can guarantee perfect recall only if the condition $\mathbf{u}_i \cdot \mathbf{u}_j = 0$, $i \neq j$, is satisfied because $\mathbf{u}_i M$ can be expressed as ([1] Eq. (17)):

$$\mathbf{u}_i M = \mathbf{u}_i \mathbf{u}_i^T \mathbf{v}_i + \sum_{j \neq i} \mathbf{u}_i \mathbf{u}_j^T \mathbf{v}_j. \tag{2}$$

This orthogonality constraint can be relaxed (at the possible expense of bidirectional recall) to simple linear independence by a more powerful recursive formulation using the Widrow-Hoff delta rule [2]:

$$M(i) = M(i-1) + \alpha \mathbf{u}_i^T (\mathbf{v}_i - \langle \mathbf{u}_i M(i-1)\rangle), \tag{3}$$

Manuscript approved July 9, 1990.

where α is a weighting factor which determines the influence of the residual error in changing the state of the memory. The attractive features of this model are that it is provably convergent when α is sufficiently small and that it yields the LMS (least mean squares) solution when the input vectors are linearly dependent. Unfortunately, like all first-order bipolar correlation memories, this model possesses the constraining symmetry that if $\langle \mathbf{u}M \rangle = \mathbf{v}$, then for the complement of $\mathbf{u}$, $\bar{\mathbf{u}}$, $\langle \bar{\mathbf{u}}M \rangle = \bar{\mathbf{v}}$. In other words, first-order bipolar AMs are incapable of storing distinct associations for complementary vectors.

THE EXPLICIT STATE REPRESENTATION

Symmetry under complementation in bipolar AMs derives from the fact that elemental correlations are computed as simple products that are invariant under commutation. This symmetry can be viewed statistically as a tacit assumption that $P(b|a) = 1 - P(b|\bar{a})$ for all elements a of a vector and all elements b of its associated vector. To relax this assumption, then, a noncommutative correlation operator is required that maps each of the four distinct elemental pair possibilities to a unique result. Fortuitously, the vector outer product operator can be employed for this purpose by simply representing $(0,1)$ states with the two-element orthonormal vectors $([0\ 1],[1\ 0])$. Vectors whose states are represented in this fashion are referred to as being in *explicit state* (ES) form. For example, the vector $[0\ 1\ 0\ 1]$ is represented in ES form as $[0\ 1\ 1\ 0\ 0\ 1\ 1\ 0]$. Although ES form is not as compact as bipolar, ES vectors can be processed exactly like bipolar vectors in the AM model described in Eq. (3) (except that the $\langle\cdot\rangle$ function uses relative magnitudes of state pairs when coercing vectors to strict ES form) in order to eliminate the complementation symmetry. Normalization issues are avoided since the ES form of every n-length binary vector has magnitude $\sqrt{n}$.

To contrast the bipolar and ES forms, consider the result of storing a mapping from training examples in which a complementary pair of vectors is associated with the same vector. In the bipolar model, this mapping of anticorrelated vectors to correlated ones results in complete destructive interference. Specifically, the sum of the two correlation matrices yields a zero matrix. In the ES model, however, no interference results. The one-to-many inverse mapping necessarily produces destructive interference for any model, but the difference between the bipolar and ES formulations is that in the bipolar model of Eq. (1), the noise equally affects the forward and inverse channels (i.e., M and M^T) while in the ES model, these channels are independent. The character of the channel independence in the ES representation is easily demonstrated given vectors $\mathbf{u}$ and $\mathbf{v}$ and observing that:

$$\begin{aligned} \mathbf{u}^T\mathbf{v} \otimes \mathbf{u}^T\bar{\mathbf{v}} &= \mathbf{0}, \\ \mathbf{u}^T\mathbf{v} \otimes \bar{\mathbf{u}}^T\mathbf{v} &= \mathbf{0}, \\ \mathbf{v}^T\mathbf{u} \otimes \bar{\mathbf{v}}^T\mathbf{u} &= \mathbf{0}, \\ \mathbf{v}^T\mathbf{u} \otimes \mathbf{v}^T\bar{\mathbf{u}} &= \mathbf{0}, \end{aligned} \tag{4}$$

where $\otimes$ is the Hadamard product $C = A \otimes B$ defined as $c_{ij} = a_{ij}b_{ij}$.

An examination of the complementation symmetry reveals that the bipolar model in Eq. (1) is only capable of learning invertible mappings. (In fact, effects from anticorrelated bit pairs introduced in the ES formulation make the bipolar form of Eq. (1) better suited for one-pass learning of invertible mappings. However, with a more sophisticated retrieval method, the ES version can surpass the bipolar performance even in this case.) The advantage of the ES representation to Eq.

(3) can be demonstrated by taking any three of the possible four 2-bit binary vectors (even the zero vector) and noting that the ES conversion transforms the linearly dependent set to one that is linearly independent. For example, the linearly dependent set $\{[0\ 1],[1\ 0],[1\ 1]\}$ is transformed to the linearly independent set $\{[0\ 1\ 1\ 0],[1\ 0\ 0\ 1],[1\ 0\ 1\ 0]\}$. A deeper analysis yields the following theorem:

Theorem: The explicit-state form of Eq. (3) converges to the following nondistributive, hence nonlinear, affine transformation:

$$\mathbf{y} = \mathbf{x}T + \mathbf{b}, \tag{5}$$

where T is a matrix and $\mathbf{b}$ is a constant vector.

Proof: Writing $\bar{\mathbf{u}}$ as $\mathbf{1} - \mathbf{u}$, and similarly for $\bar{\mathbf{v}}$, the ES forms of binary vectors $\mathbf{u}$ and $\mathbf{v}$ can be partitioned as $[\mathbf{u}|(\mathbf{1} - \mathbf{u})]$ and $[\mathbf{v}|(\mathbf{1} - \mathbf{v})]$, respectively, simply by separating the even and odd elements. The transformation then takes the form:

$$[\mathbf{u}|(\mathbf{1} - \mathbf{u})]M = [\mathbf{v}|(\mathbf{1} - \mathbf{v})]. \tag{6}$$

Partitioning M as

$$M = \begin{vmatrix} A & B \\ C & D \end{vmatrix} \tag{7}$$

yields the following bipolar expression for $\mathbf{v}$:

$$\mathbf{v}_{(bipolar)} = \mathbf{u}A + (\mathbf{1} - \mathbf{u})C - \mathbf{u}B - (\mathbf{1} - \mathbf{u})D, \tag{8}$$

which simplifies to

$$\mathbf{v}_{(bipolar)} = \mathbf{u}(A - B - C + D) + \mathbf{1}(C - D). \tag{9}$$

Letting $T = (A - B - C + D)$ and $\mathbf{b} = \mathbf{1}(C - D)$ completes the proof.

One important consequence of this result is that it assures that any mapping of a linearly independent set of input vectors can be learned even in the presence of constant additive noise. This is surprising because an additive noise vector always exists that transforms a linearly independent set of vectors to one that is linearly dependent. (A trivial case is the addition of a vector $\mathbf{z}$ to each vector in a linearly independent set S when $-\mathbf{z} \in S$.) In other words, the ES representation somewhat relaxes the linear independence condition required for perfect learning in Eq. (3). This can also be accomplished in the bipolar model by appending a constant '1' to each vector. The advantage of the ES representation is that recall from an incomplete vector (i.e., containing [0 0] states) results in a transformation that subtracts the contributions of the missing vector positions from both T and $\mathbf{b}$. In the bipolar model, however, the additive component of the affine transformation is unaffected by zero entries in the input vector. Thus, the ES representation generally should have superior recall performance from incomplete inputs than the bipolar representation.

In addition, the ES representation supports probabilistic measurements with confidence factors. Specifically, a state $[\alpha\ \beta]$ can represent knowledge that an event is true with probability α and is

false with probability β where $\alpha + \beta < 1$. For example, if a measurement process with confidence q (i.e., is accurate to within some acceptable tolerance bounds with probability q) suggests that a given event is true with probability p, this information could be encoded as $[pq \;\; (1-p)q]$. Thus, the results of a measurement having zero confidence would be treated by the recall process as though no information about the event were available. This is not equivalent to encoding the state as $[0.5 \;\; 0.5]$.

HIGHER ORDER ASSOCIATIVE MEMORIES

The definition of an AM network given in Eq. (3) uses only first-order correlations in learning a heteroassociative mapping. It is well known, however, that many important mappings (e.g., XOR) require the use of higher order correlations. Fortunately, this can be achieved without altering the AM model simply by transforming the vectors so that the higher order autocorrelation information becomes explicit in the first-order. For example, a vector $\mathbf{u}$ can be transformed to a vector $\mathbf{x}$ so that the second-order information in $\mathbf{u}$ is first-order explicit in $\mathbf{x}$ as follows:

$$\mathbf{x} \;=\; [\, \mathbf{u}_1\mathbf{u}_2 \mid \ldots \mid \mathbf{u}_i\mathbf{u}_j \mid \ldots \,] \quad i < j. \tag{10}$$

In this case the transformation is equivalent to collecting the upper triangular elements of the autocorrelation matrix $\mathbf{u}^T\mathbf{u}$ as a single vector $\mathbf{x}$. The generalization to kth-order is straightforward:

$$\mathbf{x} \;=\; [\, \mathbf{u}_1\mathbf{u}_2\ldots\mathbf{u}_k \mid \ldots \mid \mathbf{u}_{i_1}\mathbf{u}_{i_2}\ldots\mathbf{u}_{i_k} \mid \ldots \,] \quad i_1 < i_2 < \ldots < i_k. \tag{11}$$

In the bipolar representation it should be apparent that the k-term products can provide only parity information, i.e., a given product will be negative if and only if an odd number of terms are negative, otherwise it will be positive. Thus, for correlations of order > 2, it is doubtful that the information added to the bipolar representation would be of significant practical value. The direct application of this *unfolding* process to ES vectors is similarly inadequate; however, this can be remedied by generalizing the ES representation.

The motivation for the ES representation was the elimination of symmetries; however, because symmetry operations always imply information loss (in the form of irreversibility), the principle behind the ES representation can be viewed as one of information maximization. Specifically, no first-order information is lost in the vector correlation process under the ES model. Thus, the generalized ES model should perform likewise for higher order correlations. This can be accomplished by representing kth-order states with a set of orthonormal 2^k-length vectors such that any ordered k-element subset of a vector maps to a unique correlation state (in direct analog to the first-order extension of binary to ES form). Because the mapping is reversible, no loss of information occurs. Unfortunately, the transformation of a vector of length n to kth-order ES form results in a vector of length $\binom{n}{k} 2^k$, where the first factor is the binomial coefficient giving the number of k-element subsets of an n-element set. Thus, the practical use of very-high-order correlations is severely limited. However, third- and fourth-order correlations for vectors having 25 to 50 elements are within the realm of feasibility for several currently available massively parallel computers and vectors having more than 500 to 1000 elements can be processed by using second-order correlations.

The computational complexity associated with the use of high-order correlation information can be often substantially reduced by eliminating redundancy in the raw vectors. For example, in many practical applications, training vectors are generated by measuring a large number of

variables or parameters without regard to (or, possibly, without the ability to determine) their statistical independence or information content. Simply by transforming the raw data by using a Karhunen-Loeve Transform (KLT), or some similar transform that can be used to maximize information on limited channels [3,4], the effective lengths of the vectors may be dramatically reduced. Thus, the order of correlation that can be practically used may be increased.

SUMMARY

In summary, it has been shown that the commutative correlation operator used in most linear associative memory (AM) models enforces symmetries that preclude the learning of several classes of important mapping functions. It has also been shown, however, that these symmetries can be eliminated simply by using a different information representation scheme. The *explicit state* (ES) representation has been proposed as an alternative to the commonly used bipolar form and has been demonstrated to permit standard AM architectures to learn nonlinear mappings. In particular, it has been shown that the ES representation permits first-order AM models to learn nonlinear transformations of the form $\mathbf{y} = \mathbf{x}T + \mathbf{b}$. This characterization is important because its properties are directly amenable to analysis by using the known properties of affine transformations. For example, it has been noted that this transformation renders the ES formulation immune to the effects of constant additive noise. It has also been shown that the ES representation can be easily generalized for the use of higher order correlation information.

REFERENCES

1. B. Kosko, "Bidirectional Associative Memories," *IEEE Trans. on Systems, Man, and Cybernetics*, **18**:1, Jan./Feb., (1988).

2. D. E. Rummelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing, Vol. I*, MIT Press, Cambridge, MA (1986).

3. N. L. Owsley, "Adaptive Data Orthogonalization," *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, April 10-12, 109-112, Tulsa, OK (1978).

4. T. Sanger, "Optimal Unsupervised Learning in a Single-Layer Feedforward Network," *Neural Networks*, **2**, 459-473, (1989).